



PCT09

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/831,180

DATE: 01/22/2002

TIME: 16:24:19

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

P.S

ENTERED

```

4 <110> APPLICANT: Chiaki Senoo
5   Mariko Numata
7 <120> TITLE OF INVENTION: Novel Trypsin Family Serine Proteases
10 <130> FILE REFERENCE: 50026/027001
12 <140> CURRENT APPLICATION NUMBER: US 09/831,180
13 <141> CURRENT FILING DATE: 2001-05-03
15 <150> PRIOR APPLICATION NUMBER: PCT/JP99/06111
16 <151> PRIOR FILING DATE: 1999-11-02
18 <150> PRIOR APPLICATION NUMBER: JP 1998-313366
19 <151> PRIOR FILING DATE: 1998-11-04
21 <160> NUMBER OF SEQ ID NOS: 53
23 <170> SOFTWARE: FastSEQ for Windows Version 4.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 1033
27 <212> TYPE: DNA
28 <213> ORGANISM: Mus musculus
30 <220> FEATURE:
31 <221> NAME/KEY: CDS
32 <222> LOCATION: (48)...(1010)
34 <400> SEQUENCE: 1
35 cctgcctcag tgttgagct cccattgct gatgtgcagg caagccg atg aaa cga   56
36                                     Met Lys Arg
37                                     1
39 tgg aag gac aga aga aca ggc ctg ttg ctg cca ttg gtc ctc ctg ttg   104
40 Trp Lys Asp Arg Arg Thr Gly Leu Leu Leu Pro Leu Val Leu Leu Leu
41   5                               10                               15
43 ttt ggg gca tgt agc tca ctg gca tgg gta tgt ggc cgg cga atg agt   152
44 Phe Gly Ala Cys Ser Ser Leu Ala Trp Val Cys Gly Arg Arg Met Ser
45  20                               25                               30                               35
47 agc aga tcc caa caa ctt aac aat gct tct gct atc gtg gaa ggc aaa   200
48 Ser Arg Ser Gln Gln Leu Asn Asn Ala Ser Ala Ile Val Glu Gly Lys
49   40                               45                               50
51 cct gct tct gct atc gtg gga ggc aaa cct gca aac atc ttg gag ttc   248
52 Pro Ala Ser Ala Ile Val Gly Gly Lys Pro Ala Asn Ile Leu Glu Phe
53   55                               60                               65
55 ccc tgg cat gtg ggg att atg aat cat ggt agt cat ctc tgt ggg gga   296
56 Pro Trp His Val Gly Ile Met Asn His Gly Ser His Leu Cys Gly Gly
57   70                               75                               80
59 tct att ctc aat gag tgg tgg gtt cta tct gca tcc cat tgc ttc gac   344
60 Ser Ile Leu Asn Glu Trp Trp Val Leu Ser Ala Ser His Cys Phe Asp
61   85                               90                               95
63 caa cta aac aac tct aaa ttg gag atc att cat ggc act gaa gac ctc   392
64 Gln Leu Asn Asn Ser Lys Leu Glu Ile Ile His Gly Thr Glu Asp Leu

```

RAW SEQUENCE LISTING

DATE: 01/22/2002

PATENT APPLICATION: US/09/831,180

TIME: 16:24:19

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

```

65 100          105          110          115
67 agc aca aag ggc ata aag tat cag aaa gtg gac aag tta ttc ttg cac 440
68 Ser Thr Lys Gly Ile Lys Tyr Gln Lys Val Asp Lys Leu Phe Leu His
69          120          125          130
71 cca aag ttt gat gac tgg ctg ctg gac aac gac ata gct ttg ctg ttg 488
72 Pro Lys Phe Asp Asp Trp Leu Leu Asp Asn Asp Ile Ala Leu Leu Leu
73          135          140          145
75 ctg aaa tcc cca tta aac ttg agt gtc aac agg ata cct atc tgc act 536
76 Leu Lys Ser Pro Leu Asn Leu Ser Val Asn Arg Ile Pro Ile Cys Thr
77          150          155          160
79 tca gaa atc tct gac ata cag gca tgg agg aac tgc tgg gtg aca gga 584
80 Ser Glu Ile Ser Asp Ile Gln Ala Trp Arg Asn Cys Trp Val Thr Gly
81          165          170          175
83 tgg ggc att act aat act agt gaa aaa gga gtc caa ccc aca att ctg 632
84 Trp Gly Ile Thr Asn Thr Ser Glu Lys Gly Val Gln Pro Thr Ile Leu
85 180          185          190          195
87 cag gca gtc aaa gtg gat ctg tac aga tgg gat tgg tgt ggc tat att 680
88 Gln Ala Val Lys Val Asp Leu Tyr Arg Trp Asp Trp Cys Gly Tyr Ile
89          200          205          210
91 ttg tct cta tta acc aag aat atg ctg tgt gct ggg act caa gat cct 728
92 Leu Ser Leu Leu Thr Lys Asn Met Leu Cys Ala Gly Thr Gln Asp Pro
93          215          220          225
95 ggg aag gat gcc tgc cag ggc gac agt gga gga gct ctg gtt tgc aac 776
96 Gly Lys Asp Ala Cys Gln Gly Asp Ser Gly Gly Ala Leu Val Cys Asn
97          230          235          240
99 aaa aag aga aac aca gcc att tgg tac cag gtg ggc att gtc agc tgg 824
100 Lys Lys Arg Asn Thr Ala Ile Trp Tyr Gln Val Gly Ile Val Ser Trp
101          245          250          255
103 ggc atg ggc tgt ggc aag aag aat ctg cca gga gta tac acc aag gtg 872
104 Gly Met Gly Cys Gly Lys Lys Asn Leu Pro Gly Val Tyr Thr Lys Val
105 260          265          270          275
107 tca cac tat gtg agg tgg atc agc aag cag aca gcg aag gcg ggg agg 920
108 Ser His Tyr Val Arg Trp Ile Ser Lys Gln Thr Ala Lys Ala Gly Arg
109          280          285          290
111 cct tat atg tat gag cag aac tct gcg tgc cct ttg gtg ctg tct tgc 968
112 Pro Tyr Met Tyr Glu Gln Asn Ser Ala Cys Pro Leu Val Leu Ser Cys
113          295          300          305
115 cgg gct atc ttg ttc cta tat ttt gta atg ttt ctt cta acc 1010
116 Arg Ala Ile Leu Phe Leu Tyr Phe Val Met Phe Leu Leu Thr
117          310          315          320
119 tgatgattaa acgtgagact gcc 1033
121 <210> SEQ ID NO: 2
122 <211> LENGTH: 321
123 <212> TYPE: PRT
124 <213> ORGANISM: Mus musculus
126 <400> SEQUENCE: 2
127 Met Lys Arg Trp Lys Asp Arg Arg Thr Gly Leu Leu Leu Pro Leu Val
128 1 5 10 15
129 Leu Leu Leu Phe Gly Ala Cys Ser Ser Leu Ala Trp Val Cys Gly Arg

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/831,180

DATE: 01/22/2002

TIME: 16:24:19

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

```

130          20          25          30
131 Arg Met Ser Ser Arg Ser Gln Gln Leu Asn Asn Ala Ser Ala Ile Val
132          35          40          45
133 Glu Gly Lys Pro Ala Ser Ala Ile Val Gly Gly Lys Pro Ala Asn Ile
134          50          55          60
135 Leu Glu Phe Pro Trp His Val Gly Ile Met Asn His Gly Ser His Leu
136 65          70          75          80
137 Cys Gly Gly Ser Ile Leu Asn Glu Trp Trp Val Leu Ser Ala Ser His
138          85          90          95
139 Cys Phe Asp Gln Leu Asn Asn Ser Lys Leu Glu Ile Ile His Gly Thr
140          100          105          110
141 Glu Asp Leu Ser Thr Lys Gly Ile Lys Tyr Gln Lys Val Asp Lys Leu
142          115          120          125
143 Phe Leu His Pro Lys Phe Asp Asp Trp Leu Leu Asp Asn Asp Ile Ala
144          130          135          140
145 Leu Leu Leu Leu Lys Ser Pro Leu Asn Leu Ser Val Asn Arg Ile Pro
146 145          150          155          160
147 Ile Cys Thr Ser Glu Ile Ser Asp Ile Gln Ala Trp Arg Asn Cys Trp
148          165          170          175
149 Val Thr Gly Trp Gly Ile Thr Asn Thr Ser Glu Lys Gly Val Gln Pro
150          180          185          190
151 Thr Ile Leu Gln Ala Val Lys Val Asp Leu Tyr Arg Trp Asp Trp Cys
152          195          200          205
153 Gly Tyr Ile Leu Ser Leu Leu Thr Lys Asn Met Leu Cys Ala Gly Thr
154          210          215          220
155 Gln Asp Pro Gly Lys Asp Ala Cys Gln Gly Asp Ser Gly Gly Ala Leu
156 225          230          235          240
157 Val Cys Asn Lys Lys Arg Asn Thr Ala Ile Trp Tyr Gln Val Gly Ile
158          245          250          255
159 Val Ser Trp Gly Met Gly Cys Gly Lys Lys Asn Leu Pro Gly Val Tyr
160          260          265          270
161 Thr Lys Val Ser His Tyr Val Arg Trp Ile Ser Lys Gln Thr Ala Lys
162          275          280          285
163 Ala Gly Arg Pro Tyr Met Tyr Glu Gln Asn Ser Ala Cys Pro Leu Val
164          290          295          300
165 Leu Ser Cys Arg Ala Ile Leu Phe Leu Tyr Phe Val Met Phe Leu Leu
166 305          310          315          320
167 Thr

```

171 <210> SEQ ID NO: 3

172 <211> LENGTH: 1034

173 <212> TYPE: DNA

174 <213> ORGANISM: Mus musculus

176 <220> FEATURE:

177 <221> NAME/KEY: CDS

178 <222> LOCATION: (69)...(1025)

179 <223> OTHER INFORMATION:

181 <221> NAME/KEY: misc_feature

182 <222> LOCATION: 10

183 <223> OTHER INFORMATION: n = A or C or G or T/U

RAW SEQUENCE LISTING

DATE: 01/22/2002

PATENT APPLICATION: US/09/831,180

TIME: 16:24:19

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

185 <400> SEQUENCE: 3

186 **cccacgcgtn cggttgatc' aatgtgggca gggcatcaag gcaggcacca ctgcactgga** 60

187 atgacaac atg atg ctc cca ctt cta att gca ctg ctc atg gct tcc aag 110

188 Met Met Leu Pro Leu Leu Ile Ala Leu Met Ala Ser Lys

189 1 5 10

191 gga caa gct aag gac cag caa gaa tca gtt ctg tgt ggc cac aga cct 158

192 Gly Gln Ala Lys Asp Gln Gln Glu Ser Val Leu Cys Gly His Arg Pro

193 15 20 25 30

195 gcc ttc cca aac tca tca tgg ctg cca ttg cgg gag ctg ctt gag gtc 206

196 Ala Phe Pro Asn Ser Ser Trp Leu Pro Leu Arg Glu Leu Leu Glu Val

197 35 40 45

199 cag cat ggt gag ttc cca tgg caa gtg agt atc cag atg ctt ggg aaa 254

200 Gln His Gly Glu Phe Pro Trp Gln Val Ser Ile Gln Met Leu Gly Lys

201 50 55 60

203 cac ctg tgt gga ggc tcc atc atc cac cgg tgg tgg gtt ctg aca gca 302

204 His Leu Cys Gly Gly Ser Ile Ile His Arg Trp Trp Val Leu Thr Ala

205 65 70 75

207 gca cac tgc ttc ccg aga acc cta tta gaa ctg gta gca gtc aat gtc 350

208 Ala His Cys Phe Pro Arg Thr Leu Leu Glu Leu Val Ala Val Asn Val

209 80 85 90

211 act gtg gtc atg gga atc aag act ttc agt gac acc aac tta gag aga 398

212 Thr Val Val Met Gly Ile Lys Thr Phe Ser Asp Thr Asn Leu Glu Arg

213 95 100 105 110

215 aaa caa gtg cag aag atc att gct cac aga gac tac aaa ccg ccc gac 446

216 Lys Gln Val Gln Lys Ile Ile Ala His Arg Asp Tyr Lys Pro Pro Asp

217 115 120 125

219 ctt gac agc gac ctc tgc ctg ctc cta ctt gcc acg cca atc caa ttc 494

220 Leu Asp Ser Asp Leu Cys Leu Leu Leu Leu Ala Thr Pro Ile Gln Phe

221 130 135 140

223 aat aaa gac aaa atg ccc atc tgc ctg cca cag agg gag aac tcc tgg 542

224 Asn Lys Asp Lys Met Pro Ile Cys Leu Pro Gln Arg Glu Asn Ser Trp

225 145 150 155

227 gac cgg tgc tgg atg tca gag tgg gca tat act cat ggc cat ggt tca 590

228 Asp Arg Cys Trp Met Ser Glu Trp Ala Tyr Thr His Gly His Gly Ser

229 160 165 170

231 gcc aaa ggc tca aac atg cac ctg aag aag ctc agg gtg gtt cag att 638

232 Ala Lys Gly Ser Asn Met His Leu Lys Lys Leu Arg Val Val Gln Ile

233 175 180 185 190

235 agc tgg agg aca tgt gcg aag agg gtg act cag ctc tcc agg aac atg 686

236 Ser Trp Arg Thr Cys Ala Lys Arg Val Thr Gln Leu Ser Arg Asn Met

237 195 200 205

239 ctt tgt gct tgg aag gaa gtg ggc acc aac ggc aag tgc cag gga gac 734

240 Leu Cys Ala Trp Lys Glu Val Gly Thr Asn Gly Lys Cys Gln Gly Asp

241 210 215 220

243 agc ggg gca ccc atg gtc tgt gct aac tgg gag act cgg aga ctc ttt 782

244 Ser Gly Ala Pro Met Val Cys Ala Asn Trp Glu Thr Arg Arg Leu Phe

245 225 230 235

247 caa gtg ggt gtc ttc agc tgg ggc ata act tca gga tcc agg ggg agg 830

248 Gln Val Gly Val Phe Ser Trp Gly Ile Thr Ser Gly Ser Arg Gly Arg

RAW SEQUENCE LISTING

DATE: 01/22/2002

PATENT APPLICATION: US/09/831,180

TIME: 16:24:19

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

```

249      240      245      250
251 cca ggc att ttt gtg tct gtg gct cag ttt atc cca tgg atc ctg gag      878
252 Pro Gly Ile Phe Val Ser Val Ala Gln Phe Ile Pro Trp Ile Leu Glu
253 255      260      265      270
255 gag aca caa agg gag gga cga gcc ctt gcc ctc tca aag gcc tca aaa      926
256 Glu Thr Gln Arg Glu Gly Arg Ala Leu Ala Leu Ser Lys Ala Ser Lys
257      275      280      285
259 agt ctc ttg gct ggc agt cca cgc tac cat ccc ata ttg cta agc atg      974
260 Ser Leu Leu Ala Gly Ser Pro Arg Tyr His Pro Ile Leu Leu Ser Met
261      290      295      300
263 ggc tct caa ata ctg ctt gct gcc ata ttt tct gat gat aaa tca aat      1022
264 Gly Ser Gln Ile Leu Leu Ala Ala Ile Phe Ser Asp Asp Lys Ser Asn
265      305      310      315
267 tgc taagctctg      1034
268 Cys
272 <210> SEQ ID NO: 4
273 <211> LENGTH: 319
274 <212> TYPE: PRT
275 <213> ORGANISM: Mus musculus
277 <400> SEQUENCE: 4
278 Met Met Leu Pro Leu Leu Ile Ala Leu Leu Met Ala Ser Lys Gly Gln
279 1      5      10      15
280 Ala Lys Asp Gln Gln Glu Ser Val Leu Cys Gly His Arg Pro Ala Phe
281      20      25      30
282 Pro Asn Ser Ser Trp Leu Pro Leu Arg Glu Leu Leu Glu Val Gln His
283      35      40      45
284 Gly Glu Phe Pro Trp Gln Val Ser Ile Gln Met Leu Gly Lys His Leu
285      50      55      60
286 Cys Gly Gly Ser Ile Ile His Arg Trp Trp Val Leu Thr Ala Ala His
287 65      70      75      80
288 Cys Phe Pro Arg Thr Leu Leu Glu Leu Val Ala Val Asn Val Thr Val
289      85      90      95
290 Val Met Gly Ile Lys Thr Phe Ser Asp Thr Asn Leu Glu Arg Lys Gln
291      100      105      110
292 Val Gln Lys Ile Ile Ala His Arg Asp Tyr Lys Pro Pro Asp Leu Asp
293      115      120      125
294 Ser Asp Leu Cys Leu Leu Leu Ala Thr Pro Ile Gln Phe Asn Lys
295      130      135      140
296 Asp Lys Met Pro Ile Cys Leu Pro Gln Arg Glu Asn Ser Trp Asp Arg
297 145      150      155      160
298 Cys Trp Met Ser Glu Trp Ala Tyr Thr His Gly His Gly Ser Ala Lys
299      165      170      175
300 Gly Ser Asn Met His Leu Lys Lys Leu Arg Val Val Gln Ile Ser Trp
301      180      185      190
302 Arg Thr Cys Ala Lys Arg Val Thr Gln Leu Ser Arg Asn Met Leu Cys
303      195      200      205
304 Ala Trp Lys Glu Val Gly Thr Asn Gly Lys Cys Gln Gly Asp Ser Gly
305      210      215      220
306 Ala Pro Met Val Cys Ala Asn Trp Glu Thr Arg Arg Leu Phe Gln Val

```

Use of n and/or Xaa has been detected in the Sequence Listing.
 Review the Sequence Listing to insure a corresponding
 explanation is presented in the <220> to <223> fields of
 each sequence using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/831,180

DATE: 01/22/2002

TIME: 16:24:20

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: N:\CRF3\01182002\I831180.raw

L:186 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:1237 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:47
L:1273 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:49